Simple Market Protocols for Efficient Risk Sharing*

Marco LiCalzi
<licalzi@unive.it>
Dept. of Applied Mathematics
University of Venice

Paolo Pellizzari
<paolop@unive.it>
Dept. of Applied Mathematics
University of Venice

(April 2005)

Abstract. This paper studies the performance of four market protocols with regard to allocative efficiency and other performance criteria such as volume or volatility. We examine batch auctions, continuous double auctions, specialist dealerships, and a hybrid of these last two. All protocols are practically implementable because the space of messages for traders is simple. We test the protocols by running (computerized) experiments in an environment that controls for traders' behavior and rules out any informational effect. We find that all protocols generically converge to the efficient allocation in finite time. An extended comparison over other performance criteria produces no clear winner, but the presence of a specialist is clearly associated with the best all-round performance.

Keywords: market microstructure, allocative efficiency, comparison of market institutions, performance criteria.

JEL Classification Numbers: G19, D61, D44, C63.

Correspondence to:

Marco LiCalzi	Dept. of Applied Mathematics, University of Venice
	Dorsoduro 3825/e
	30123 Venezia, Italy
Phone:	[++39] (041)-234-6925
Fax:	[++39] (041)-522-1756
E-mail:	licalzi@unive.it

* We thank Alessandro Sbuelz and audiences at Bocconi and Catholic University of Milano for their comments. Financial support from MIUR is acknowledged.

1 Introduction

Financial markets where agents exchange risky assets serve two main purposes. First, they allocate risk among traders and improve allocative efficiency. Second, they diffuse traders' private information and facilitate information diffusion. The simultaneous pursuit of allocative and informational efficiency is usually impossible. Different market arrangements are more favorable to the search for different notions of efficiency. We observe that the state of knowledge in this respect is remarkably unbalanced.

There is a vast literature on market microstructure that is especially keen on the analysis of the conditions affecting the revelation (and exploitation) of private information. On the other hand, much less attention has been devoted to the functioning of financial markets with respect to allocative efficiency. This problem is the focus of our paper, which aims to provide the experimental evidence needed to ground a theoretical analysis.

We study the performance of four market protocols with regard to allocative efficiency and other performance criteria such as volume or volatility. These additional criteria are usually extolled by exchange regulators because they can be objectively measured and provide useful proxies for the evaluation of a market protocol. The four market protocols that we examine are: the batch auction, the continuous double auction, a special form of (nondiscretionary) specialist dealership, and a hybrid of these last two. Contrary to theoretical constructs such as Walrasian tâtonnement, these four protocols are practically implementable because the space of messages that traders need to post is simple.

We test the protocols by running (computerized) experiments in an environment that controls for traders' behavior and rules out any informational effect. The behavior of the agents span how they formulate trading strategies, how they form expectations, and how they interpret signals. Working with agent-based simulations instead of human agents permits to isolate the impact of the trading protocols from these behavioral components. In standard laboratory experiments, instead, it is not possible to extricate the interactions between protocol and behavioral effects.

The main behavioral limitation on our agents is that they exhibit limited intelligence, similarly to the "zero intelligence" traders in Gode and Sunder (1993). The price of the risky asset is the main driver for their choices; but, like real traders in real markets, they ignore the correct equilibrium price and thus lack an essential piece of information to compute the efficient allocation. This leaves the market protocol in charge of "discovering" the right price for them. From a roaring and confused crowd of traders each trying to (guess and) achieve his preferred risk allocation, the market protocol must extract and send out price signals that point traders in the right direction. This makes convergence to the "right" price a necessary condition for allocative efficiency. Assuming that there is sufficient liquidity in the market, we find that all protocols generically converge to the efficient allocation and to the equilibrium price in finite time.

We then turn to a dynamic analysis of the performances. For practical purposes, it is probably more important to know how protocols perform during the (perhaps long) transient period before they achieve the efficient allocation. We study how long it takes for different protocols to discover the equilibrium price and how fast they lead traders to the efficient risk sharing allocation. We measure the volume of trade developed to reach the efficient allocation from the initial endowment, as well as the volatility of the time series of prices. This extended comparison over several dynamic performance criteria produces no clear winner, but the presence of a nondiscretionary specialist dealer is clearly associated with the best all-round performance. Continuing the analogy above, the introduction of an nondiscretionary "center" in a market protocol seems to improve its ability to stabilize and direct traders' own groping for the right price.

The organization of the paper is the following. Section 2 describes the model tested in our experiments. In particular, Section 2.5 provides a comparative review of our model against the relevant literature. Section 3 details the experimental design and provides detailed instructions for its replication. Section 4 reports on the results obtained and Section 5 offers our conclusions.

2 The model

Following Smith (1982), we identify three distinct components for our (simulated) exchange markets. The environment in Section 2.1 describes the general characteristics of the economy, including agents' preferences and endowments. The market protocols in Section 2.2 provide the institutional details which regulate the functioning of an exchange. The behavioral assumptions in Section 2.3 specify how agents make decisions and take actions. Section 2.4 details a few alternative behavioral assumptions that have been used to test the robustness of our conclusions. Finally, Section 2.5 compares how our assumptions fares with those prevailing in the received literature.

2.1 The environment

We consider a two-asset economy with n traders. The two available assets are a risky stock and cash. The rate of interest is normalized to zero, so cash acts as the numeraire. The stock pays no dividends and has a (random) realization value Y at a given time T in the far future. Each trader i has an initial endowment of cash $c_i \geq 0$ and shares $s_i \geq 0$. The total amount of cash and stock in the economy is $C = \sum_i c_i > 0$ and $S = \sum_i s_i > 0$.

To rule out any informational effect, we assume that all traders believe that Y is normally distributed with mean $\mu \geq 0$ and precision $\tau = 1/\sigma^2 > 0$ and that no new information is ever released. Therefore, traders' beliefs about Y are homogeneous and never change until uncertainty resolves.

Each trader i has "cara" preferences over his final wealth, with a coefficient of risk tolerance $k_i > 0$. Therefore, trader i's excess demand function for stock (net of his endowment s_i) is the linear function

$$q_i(p) = k_i \tau(\mu - p) - s_i. \tag{1}$$

Let $K = \sum_i k_i$ be the sum of traders' coefficients of risk tolerance. By well-known results pioneered in Wilson (1968), the unique efficient risk-sharing allocation for this economy requires that trader i holds $s_i^* = (S/K)k_i$ shares of the stock. In other words, the efficient allocation is unique and proportional to the coefficient of risk tolerance.

2.2 The market protocols

Clearly, the competitive equilibrium achieves the efficient allocation in this environment. The zero aggregate excess demand condition implies

$$p^* = \mu - \frac{S}{\tau K}. \tag{2}$$

At price p^*, the trader i's net demand

$$q_i(p^*) = \left(\frac{S}{K}\right) k_i - s_i$$

is exactly filled, making his final allocation $q_i(p^*) + s_i$ equal to the required $s_i^* = (S/K)k_i$. Hence, if a market protocol attains the competitive equilibrium, it implements the efficient allocation.

The issue, however, is that the informational requirements for a competitive equilibrium are often not realistic. For instance, the fictitious protocol of the Walrasian auctioneer requires an iterative process during which traders communicate their entire excess demand function to a centralized market maker before any trade takes actually place. Realistic market protocols are much simpler, in the sense that they require far less information from traders.

We compare the performances of four *simple* market protocols: a batch auction, a continuous double auction, an automated dealership, and a hybrid market. The first protocol is simultaneous, while the other three are sequential. Except where otherwise noted, the following features are common to all the four protocols.

A protocol is organized in trading sessions (or days). Agents participate to every trading session, but each of them can exchange at most one unit in each trading day. During a trading session, an agent can buy or sell at most one unit of the risky asset. If the protocol is sequential, the order in which agents place their orders is randomly chosen for each trading session. If the protocol is simultaneous, all order are made known and processed simultaneously so the time of their submission is irrelevant. In every trading session, each agent selects randomly one side of the market where he attempts to place a trade: he can switch roles across trading sessions, but he cannot place simultaneous orders for buying and selling within the same session. The books are completely cleared at the end of each trading session.

Prices are quoted using a minimum tick; in other words, they are discretized. Moreover, prices must be nonnegative: if a trader places a bid lower than zero, this is ignored; if a trader places an ask lower than zero, this is automatically converted to the lowest strictly positive price compatible with the existing tick.

Batch auction. In each trading session, after traders submit their orders, the aggregate excess demand function is computed and the exchange price p^* is determined by setting the aggregate excess demand equal to zero. If there are multiple solutions, we select the midpoint of the interval between the lowest and the highest clearing price. (If there are no solutions, no exchange takes place.) Shares and corresponding payments are exchanged between traders who submitted bids not lower than p^* and asks not higher than p^*. Traders who placed orders exactly at price p^* may be accordingly rationed. This protocol is also known as the k-double auction, with $k = 1/2$.

Continuous double auction. In each trading session, traders place their orders on the selling and buying books. Their orders are immediately executed if they are marketable; otherwise, they are recorded on the books with the usual price-time priority. Orders are canceled only when a matching order arrives or the trading day is over.

Automated dealership. There is a specialist dealer who posts bids and asks valid only for a unit transaction. Agents check sequentially the dealer's quotes for the side of the transaction they are attempting. If an agent accepts the dealer's quote, the exchange takes place at the quoted price. Right after a transaction is completed, the two dealer's quotes for bid and ask increase by one if the agent completed a purchase and decrease by one otherwise. The size of the bid-ask spread stays fixed over time, so the price is never unique. Limited to this protocol, therefore, convergence of prices to a given value p^* should be interpreted as convergence to prices to a bid-ask interval that contains p^*.

Hybrid market. This protocol combines the continuous double auction with the automated dealership. Distinct selling and buying books hold quotes from the specialist dealer and from the public, respectively. The dealer must post bids and asks valid only for a unit transaction and revises her quotes as in the automated dealership; in particular, she moves her quotes only after transactions in which she has been involved. Agents check sequentially the books for the side of the transaction they are attempting. Their orders are immediately executed at the best price available (which may be different from the specialist's) if they are marketable; otherwise, they are recorded on the traders' book with the usual price-time priority. Agents' orders are canceled only when a matching order arrives or the trading day is over. Hence, once deposited on the traders' book, an order from an agent cannot be executed with the dealer.

We note two limitations to the realism of our assumptions about the market protocols. First, in a sequential protocol, the order of arrival for agents is randomly chosen: this mutes any issue concerning the trade-off between efficacy and immediacy. Second, we assume that agents can trade at most one unit per trading session: this circumvents the problem of choosing the order size. Similar restrictive assumptions are common in the literature; see for instance Glosten and Milgrom (1985), which has inspired our version of specialist dealership.

2.3 Behavioral assumptions

A major obstacle in the study of microeconomic systems is that their performance is jointly determined by the interactions of traders' behavior within the market protocol. As traders may react differently to different market protocols, it is difficult to separate the intrinsic characteristics of a market protocol from the properties induced by the traders' strategies. We follow the usual approach and concentrate on the institutional characteristics of the protocols, by making general-purpose assumptions on traders' behavior that constrain their freedom of choice. Except where otherwise noted, these assumptions hold for all the (computerized) experiments reported in this paper.

There are three established assumptions in the literature. One is that traders are restricted to trade one unit at a time. This restriction on traded quantities simplifies the strategy space and allows direct comparisons with existing theoretical results. The second assumption states that buying orders are constrained by the available cash and selling orders by the available endowment of stock; that is, budget constraints hold. This is consistent with a value-based strategy ("buy low, sell high"), which is a seemingly natural requirement of rationality for traders' behaviors. The third assumption is that each trader has a constant valuation for each unit traded. We maintain the first two assumptions, but relax the third one.

In our setting, the demand function (1) of each trader is strictly decreasing. The assumption of constant unit valuations is naturally generalized by deducing the valuation for the next unit to trade from the demand function. If the current endowment of a trader is s_i, we invert his unit excess demand function $\pm 1 = k_i \tau(\mu - p) - s_i$ and derive his valuation for the next unit to trade as

$$p(\pm 1) = \mu - \frac{s_i \pm 1}{k_i \tau}, \tag{3}$$

where the $\pm$ sign depends on whether the attempted trade is a purchase or a sale. Clearly, this implies that the reservation price of each trader depends on the side of the transaction he is entering and on his current endowment s_i; moreover, his (implicit) bid-ask spread is $2/(k_i \tau)$.

Given his valuation, a trader must decide which side of the transaction he wants to attempt and (if necessary) what price to offer. We assume that, at the start of a trading session, each trader chooses either side with equal probability. This randomized choice is stochastically independent of previous history, endowment, or any other parameter of the model. After the choice of the trading side is made, therefore, a trader "knows" that he is going to be a buyer (or a seller) and that his valuation for the next unit he will attempt to buy (or sell) is $p(+1)$ (or $p(-1)$) from Equation 3. Once his trading intentions are known, a trader (deterministically) places a bid or an ask equal to his valuation. We nickname this assumption by TT, as a mnemonic for "truth-telling" behavior.

The common properties of our market protocols may impose two deviations from truth-telling. First, the valuation of a trader may be a number different from the ticked prices. As detailed below, we adopt an "exact" experimental design that rules out this case. Second, the valuation of a trader may be negative; in this occurrence, all protocols refuse negative bids and automatically update a negative ask to the lowest non-zero ticked price.

2.4 Alternative behavioral assumptions

The set of assumptions in Section 2.3 uniquely determines traders' behavior in each of the different market protocols examined in this paper. In our simplified environment, therefore, the only differences in performances are due only to the institutional differences embodied in the market protocols. This insulation from spurious effects (due to traders' behavior) makes it possible to evaluate market protocols on their own. Clearly, this insulation carries a cost in realism because it is a sensible assumption that real traders adjust their strategies to the type of market protocol they are forced to use. Rather, the agents in our simulations exhibit zero intelligence under several respects. They do not react to differences in the market protocol; they do not attempt to trade strategically; they do not update their behavior rules over time.

Given that the purpose of the study is to understand the performance of market protocols *per se*, we justify the assumption of zero intelligence because the experimental design should try as much as possible to keep traders' behavior unchanged across different institutions. On the other hand, since there are several ways to achieve this objective, we have tested the robustness of our conclusions under two different sets of behavioral assumptions. Compared to how TT solves the tradeoff between efficacy and immediacy, the first behavioral assumption sacrifices some immediacy for greater efficacy while the second one goes in the opposite direction. This allows us to rank the behavioral assumptions used in this paper from least to most efficacious or, equivalently, from most to least immediate. Hereafter is the description of these two alternative sets of behavioral assumptions.

The first set of assumptions is directly inspired by Gode and Sunder (1993). Under the heading of "zero intelligence", they assume that buyers bid a price uniformly drawn from the interval $[0, p(+1)]$ and sellers ask a price uniformly drawn from the interval $[p(-1), 2p^*]$, where p^* is the competitive equilibrium price. We nickname their implementation as GS. Compared to TT, the GS assumption is a closer approximation to a common form of strategic trading that misrepresent valuations in order to extract more surplus from a transaction. This makes it harder to complete a transaction and thus reduces immediacy; on the other hand, conditional on trading taking place, it increases the probability of a better price and thus improves efficacy. Clearly, the GS assumption also introduces an additional source of noise in the model which may confound the results.

The second set of assumptions is inspired by a notion of mental accounting similar to the arguments in Thaler and Johnson (1990). They report experiments showing that gambling behavior is affected by the sign of the current net monetary position with respect to the initial one. People tend to increase their propension to risk in the presence of a prior gain. In our context, this suggests that the traders should increase their willingness to trade when their past transactions have been more profitable.

We implement this requirement modifying the valuation price in TT as follows. Consider a trader i who has a valuation v for one unit and buys it at price of p, making a "gain" of $v - p$; similarly, let $p - v$ the gain of the trader when he completes a sale. At each trading session t, given the (possibly empty) history of his past trades before t, a trader has matured a trading surplus G equal to the (undiscounted) sum of his past gains. We assume that, at time t, the valuation of a buyer is his price $p(+1)$ from (3) increased by his current trading surplus G. Similarly, the valuation of a seller is $p(-1) - G$. We nickname this implementation MA as a mnemonic for mental accounting.

Compared to TT, the MA assumption makes traders who have had a lucky streak of transaction prices in the past more willing to trade; on the other hand, it does not affect the behavior of the "unlucky" players who had transacted exactly at their valuation price. Clearly, this makes it generically easier to complete a transaction and thus improves immediacy; moreover, conditional on trading taking place, it increases the probability of a worse price and thus reduces efficacy. The MA assumption does not introduce additional sources of noise in the model.

2.5 Comparison with the literature

The seminal contribution about computerized experiments in a controlled environment is Gode and Sunder (1993). This paper establishes the allocative efficiency of the continuous double auction under zero intelligence or, in our jargon, under the GS assumption. Our analysis extends this work into two directions. First, we explicitly compare the allocative efficiency of several different market protocols and find that they all share the ability to lead zero intelligence traders towards the efficient allocation. Second, we replace their assumption of constant unit valuations for each trader with decreasing unit valuations. In particular, this would dispose of the critique in Cliff and Bruten (1997) about possible nonconvergence to the equilibrium price.

Bottazzi, Dosi and Rebesco (2005) compare the allocative efficiency and other performance criteria for three protocols (Walrasian tâtonnement, the batch auction, and the continuous double auction) under behavioral assumptions very different from ours. They consider an ecology of trend followers and noise traders and they allow traders to place variable-size orders which are generated using an *ad hoc* rule. Their study is unable to separate behavioral from market effects, leading to the somewhat weak conclusion reported in the abstract: "The results highlight the importance of the institutional setting in shaping the dynamics of the market but also suggest that the latter can become the outcome of a complicated interaction between the trading protocol and the ecology of traders' behaviors. In particular, we show that market architectures bear a central influence upon the time series properties of market dynamics. Conversely, the revealed allocative efficiency of different market settings is strongly influenced by the trading behavior of agents."

Audet, Gravelle, and Young (2002) compare execution quality in a batch auction and in a competitive (and discretionary) dealership market. The paper argues that execution quality is a multiattribute concept, whose measurement involves some combination of different measures of performance. From a behavioral point of view, the paper assumes liquidity traders subject to informational shocks and derives choices using a complex neural network approximation to Nash behavior. The main conclusion is that (their own variant of) the competitive dealership provides superior execution quality when trading is thin and correlated or when there are large liquidity shocks.

Satterthwaite and Williams (2002) view a market protocol as an algorithm to solve the problem of allocative efficiency. Agents have constant unit valuations and are restricted to trade at most one unit per trading session. They assume that agents view trading as a repeated one-shot game with incomplete information and play equilibrium strategies. Consequently, MA behavior is ruled out. Under these assumptions, the paper proves that the batch auction is worst-case asymptotically optimal: as the number of agents grows, it forces the worst-case inefficiency to zero at the fastest possible rate. Transient effects or other measure of performance are not considered.

3 Experimental design

3.1 Identification

A simulation run for our model requires the specification of five global parameters, a list of individual variables for each trader, as well as specific assumptions about market protocol and traders' behavior. The global parameters are the number n of traders, the mean μ and the variance σ^2 of the realization value Y of the stock, the number t of trading sessions, and the size Δ of the tick. Individually, a trader i is characterized by his coefficient k_i of risk tolerance and by his endowment of cash c_i and stock s_i. Finally, for protocols involving the dealer, we need to select her initial quotes.

The market protocols are described in Section 2.2. For ease of reference, we nickname these four protocols as B (batch auction), C (continuous double auction), D (automated dealership), and H (hybrid market). Similarly, our three sets of behavioral assumptions are described in Section 2.3 and 2.4. They have been nicknamed TT (truth-telling), GS (Gode and Sunder), and MA (mental accounting).

We have run (computerized) simulations for all $4 \times 3 = 12$ possible combinations of protocols and behavioral assumptions, over different instantiations of the parameters. The results reported in Section 4 seem to be robust to substantial changes in the parameters. The only exception that needs a separate study (left to future research) is the case where the overall liquidity $C = \sum_i c_i$ of the system is very low. Therefore, to simplify the presentation, we fix the exemplar parametric configuration reported in Table 1 and report the simulations for the four market protocols under different behavioral assumptions. The initial dealer's quotes are a bid of 745 and an ask of 751, with a fixed bid-ask spread of 6.

	Parameters		Initialization
Global	n	$=$	$1,000$
	μ	$=$	$1,000$
	σ^2	$=$	120
	t	$=$	2,500
	Δ	$=$	1
Trader	k_i	$=$	divisors of σ^2 in $[10, 40]$
	c_i	$=$	$50,000$
	s_i	$=$	permutation of $2k_i$

Table 1: Exemplar for identification.

While the choice of the parameters is largely irrelevant, there are some features that deserve to be noted. Assume TT and recall that buying and selling valuations for one unit obey (3). Since the tick is $\Delta = 1$, ticked prices must be integers. If the right-hand side of (3) is always an integer, traders have integer valuations and the market protocol need not round bids and asks to a ticked price. In this case, we say that the simulation *exactly implements* TT because the messages passed by agents are never modified (even minimally) by the market protocol. The most important consequence of an exact implementation is that it allows *exact* convergence to the equilibrium price supporting the efficient allocation. (Properly speaking, exact convergence is not relevant under the automated dealership protocol, because the fixed bid-ask spread prevents the price from being unique.)

We have taken care to ensure that our exemplar case exactly implements the three market protocols under study. To this purpose, we choose integer values for μ and σ^2 and initialize each k_i's with a stochastically independent draw from a uniform distribution over a subset of the divisors of σ^2. (The choice $\sigma^2 = 120$ makes sure that there is adequate variety among the divisors.) In our highly structured context, this is sufficient for exact implementation under both TT and MA, as well as under GS if one assumes (as we do) that the support of the uniform draws over bids and asks is formed by the integers in the two characterizing intervals.

The appeal of exact implementation is mostly theoretical: we do not expect real market protocols to be sufficiently fine to make exact implementation possible in practice. Moreover, the impact of approximate implementation is likely to be minimal and, consequently, most simulations in the literature do not bother at all with it. On the other hand, we believe that computerized simulations should explicitly deal with this issue for two reasons. First, the experimental testing of market protocols should approximate as closely as possible ideal conditions: exact implementation is necessary for convergence to allocative efficiency, which is probably the most important performance criterion. Second, as part of the experimental study, one might also try and assess how the failure of exact implementation affects the performance of a protocol. For the purpose of this study, it suffices to note that all the simulations reported satisfy the requirement of exact implementation.

A second and minor feature is that the parameters in the exemplar are chosen to simplify the computation of the competitive equilibrium price given by (2). We initialize the vector of s_i's by taking a random permutation of the vector of $2k_i$'s. This implies that $S = \sum_i s_i = 2\sum_i k_i = 2K$ and therefore, by (2), the competitive equilibrium price is $p^* = \mu - 2\sigma^2 = 760$ in all of our reported simulations.

Figure 1 provides a visual summary of a typical instantiation of the coefficients of risk tolerance k_i's and of the initial endowment s_i's. The coefficients of risk tolerance are always divisors of the variance, and the initial endowments are random (and stochastically independent) permutations of these. Therefore, the marginal frequency distribution of the coefficients of risk tolerance is taken to be independent of the initial endowments.

Finally, we stress that the parameters in the exemplar are not an attempt to calibrate the model to any specific set of real data. We do not claim that our model has the descriptive power that would warrant a calibration exercise. Our simulations are blind to any informational effects and thus are not fit to replicate the price dynamics observed in real markets. The purpose of this study is limited to gather evidence on the performance of different market protocols with regard to allocative efficiency, and our results do not extrapolate to markets subject to informational effects.

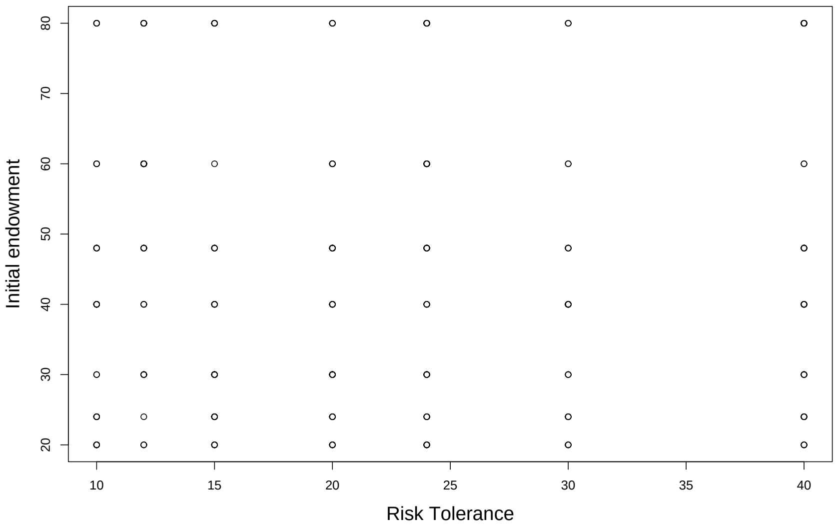

Figure 1: Initial endowments versus risk tolerances.

3.2 The simulations

A round of testing requires to simulate $4 \times 3 = 12$ possible combinations of protocols and behavioral assumptions. In order to keep experimental conditions as comparable as possible, a typical round of simulations runs as follows. For each batch of 12 combinations, we instantiate parameters according to the same exemplar. We have tested a large range of exemplars. For simplicity, all simulations reported in this paper share the initialization reported in Table 1.

We also try to reduce to a minimum the impact of randomness on the simulations. Under TT and MA, the only sources of randomness are in the order in which agents are sampled under a sequential market protocol and in the choice of which side of the transaction they attempt to complete. Under GS, a third source of randomness is in the selection of bids and asks from two intervals of possible choices. For each batch of simulations, we use the same (randomly chosen) sampling and the same (randomly chosen) selection of transacting sides. In other words, the sequence in which players place orders and the side they attempt to transact is the same within each batch of simulations.

At the end of a batch of simulations, we record the time series for prices, volume, and endowments, and compute relevant statistics for the performance criteria discussed in the next section. The simulations have been run using a dedicated package of routines written in Pascal.[1]

3.3 The performance criteria

There are several criteria that can be used to evaluate the performance of a market protocol. One of our two focuses of interest is the ability of market protocols to attain the efficient allocation, so we report on their convergence to the efficient allocation, as well as on the speed with which they achieve it. We measure traded volume, because higher volumes signal less effective protocols that let unnecessary trades take place. The second focus of interest is the dynamic behavior of a protocol ("how does it get there?") and therefore we also keep track of the evolution of these and other appropriate indicators over time. For instance, we report standard deviations and kurtosis for the time series of the closing prices. These are useful indicators for assessing the stability of prices over time, even though Stigler (1964) has long ago warned that one cannot take for granted that "smoothness of price movements is the sign of an efficient market" (p. 125). A detailed description of the performance criteria used follows hereafter.

The first basic criterion of allocative efficiency is whether a protocol converges to the efficient allocation in finite time. Convergence to the efficient allocation is generically achieved, at a price equal to the competitive equilibrium price p^*. Clearly, in any of the four protocols the current price may hit p^* before the efficient allocation is attained. Therefore, we evaluate the speed of convergence by recording the number of trading sessions completed before no further trading takes place. For short, we call this number NT as a mnemonic for "no trade". To evaluate the dynamic properties of the protocols, we also keep track over time of the distance of the current allocation from the efficient one.

Volume is measured by the total number of one-unit transactions completed before attaining convergence to the efficient allocation. We also monitor volume over time, measured by the cumulative number of transactions completed within the first t trading sessions. Under the automated dealership protocol, the transfer of one unit from a trader to another one must go through the dealer and therefore requires two transactions, instead of just one. Whenever two matching transactions go through the dealer, we record them as one so that the statistics for volume report for each protocol the effective number of transactions completed between agents. This makes volume directly comparable across protocols, even if it fails to record the few units of unmatched inventory that may remain with the specialist.

Another evaluation criterion is the subjective welfare of the traders, which we measure

[1]The source code for the exemplar is available at `http://venus.unive.it/licalzi/research.html`.

by the certainty equivalent of their current position. The advantage of using certainty equivalents over utilities is that we can compute (and compare) the monetary value of an allocation by summing up the certainty equivalents of all traders. We note that protocols such as the batch auction or the continuous double auction are self-contained, in the sense that traders exchange cash and stock only among themselves. The overall value of an efficient allocation is of course always the same. However, when trading among agents goes through a dealer, some money is lost because of the bid-ask spread. Therefore, we expect the overall value of the allocation to be lower for protocols which are not self-contained.

Finally, we keep track of the standard deviations and the kurtosis of the time series of closing prices. (The closing price is the last price at which a transaction has taken place by the end of the current trading session.) For each simulation, we compute these two statistics over the closing prices between the first trading session and the last active trading session; that is, from time $t = 1$ to NT. As usual, we report the standardized kurtosis κ; that is, $\kappa = \mu_4/\mu_2^2 - 3$, where μ_i is the i-th central moment of the empirical distribution. It is well-known that $\kappa = 0$ for a normal distribution. From a dynamic point of view, we report the standard deviations over closing prices computed over a moving time window formed by the last 20 trading sessions.

4 Results

For each of the four market protocols, we have run several simulations under the different three behavioral assumptions. As discussed below, replacing TT with GS implies only minor changes that we compactly summarize in Section 4.2. To simplify the presentation, we fix the choice of parameters as described above in Table 1 and study the variations principally across the four protocols for the TT and the MA behavioral assumptions. A separate robustness analysis discusses the choice of different parameters.

As discussed in Section 2.1, the efficient allocation is unique and proportional to the coefficient of risk tolerance. Given our exemplar, this corresponds to each trader holding a final endowment of stock equal to twice his coefficient of risk tolerance. A visual summary of the efficient allocation is in Figure 2. Assuming sufficient liquidity (as we do in this paper), all simulations converge to this efficient allocation in finite time. In particular, for the exemplar discussed here, the implementation is exact and the price converges to the competitive equilibrium price $p^* = 760$.

From a static point of view, the obvious benchmark for the comparative evaluation of our performance criteria is the Walrasian allocation mechanism known as tâtonnement. While

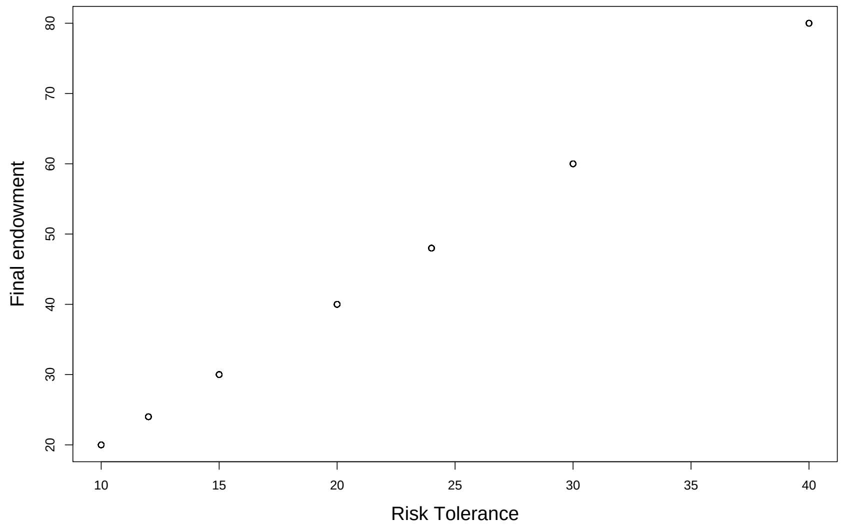

Figure 2: Final endowments versus risk tolerances.

its informational requirements are unwieldy and thus its practical interest is very limited, in our context the tâtonnement protocol is guaranteed to yield the efficient allocation in one (giant) step. This protocol also attains the competitive equilibrium price p^*. Finally, and perhaps more interestingly, it minimizes the number of transactions needed to achieve the efficient allocation because it correctly matches traders with positive excess demand with agents with negative excess demand.

4.1 The TT assumption

Table 2 reports summary statistics computed as averages over 25 different batches of simulations under the TT behavioral assumption. We report the averages because they are more robust to individual variations, but we also note that no significant departures from these statistics were observed in the individual simulations. Within each batch, we use the same (randomized) choices over the four protocols, as explained in Section 3.2. The table reports several pieces of information for each of the four protocols, as listed in the first column. We comment on each of them.

The second and third columns give the traded volume (Vol) and the percentage of excess volume (ExcV) with respect to the number of transactions that the Walrasian protocol would require to achieve in one step the efficient allocation. The batch auction and the specialist dealership generate minimal excess volume: which one performs better in this

Prot.	Vol	ExcV	NT	CE	Loss	Dist	SD	Kurt
B	11322.44	2.7%	148.32	88079.50	0.00%	0.00	4.57	6.288
C	28778.04	161%	149.40	88079.50	0.00%	0.00	31.72	3.405
D	11157.48	1.6%	143.56	88022.26	0.065%	0.11	4.09	0.754
H	16221.56	47%	144.88	88050.15	0.033%	0.01	2.77	-0.020

Table 2: TT: summary statistics for 25 passes with $n = 1000$.

respect depends on how far are the initial dealer's quotes from the equilibrium price. The continuous double auction is seriously wasteful, while the hybrid protocol sits in between its parents (dealership and continuous double auction). The ranking with respect to volume is $\{B, D\} > H > C$, where $>$ stands for "less volume" and the notation $\{B, D\}$ means that the ranking is not conclusive. (This applies when our extensive analysis has found reversals in ranking across simulations, usually different from the exemplar.) From a dynamic point of view, Figure 3 confirms this ranking. (B is barely visible because it is covered by D.)

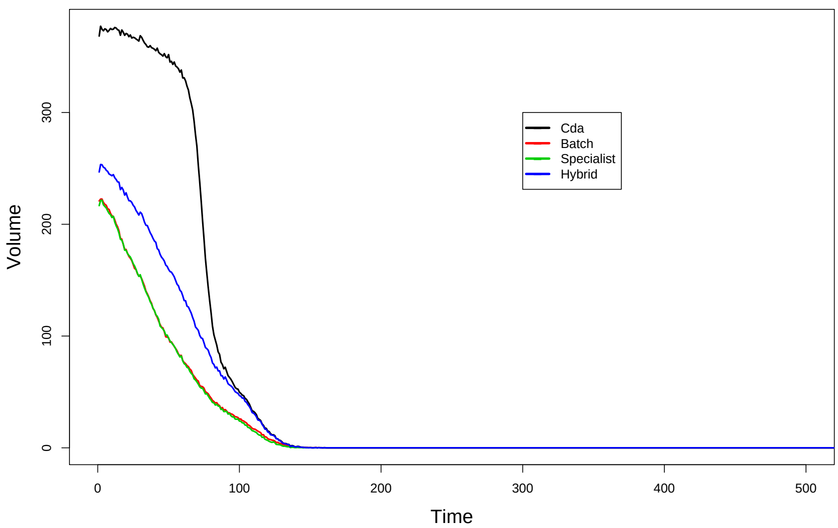

Figure 3: TT: volume versus time.

The fourth column in Table 2 computes the number (NT) of trading days necessary to achieve no trade and hence the efficient allocation. In our exemplar case, the maximum number of units between the initial endowment and the final efficient endowment is 60, so this is a lower bound on the number of trading sessions required to achieve allocative efficiency. The seventh colum reports the distance (Dist) of the final allocation from the efficient allocation. (We measure the distance using the norm one normalized by the number of traders: thus, $d(e, e^*) = \sum_i |e_i - e_i^*|/n$.) All protocols achieve the efficient allocation, sometimes up to one unit for one trader who cannot find the counterpart for his last unit trade. The ranking with respect to convergence is $D > H > B > C$, where $>$ stands for "smaller time to no trade". The differences are very small, but persistent.

From a dynamic point of view, the ability of all protocols to achieve allocative efficiency is shown in Figure 4, which reports the distance of the current allocation from the efficient one with respect to the cumulative volume. (B is barely visible because it is covered by D.) This graph provides an implicit indicator of the ability of a protocol to minimize the number

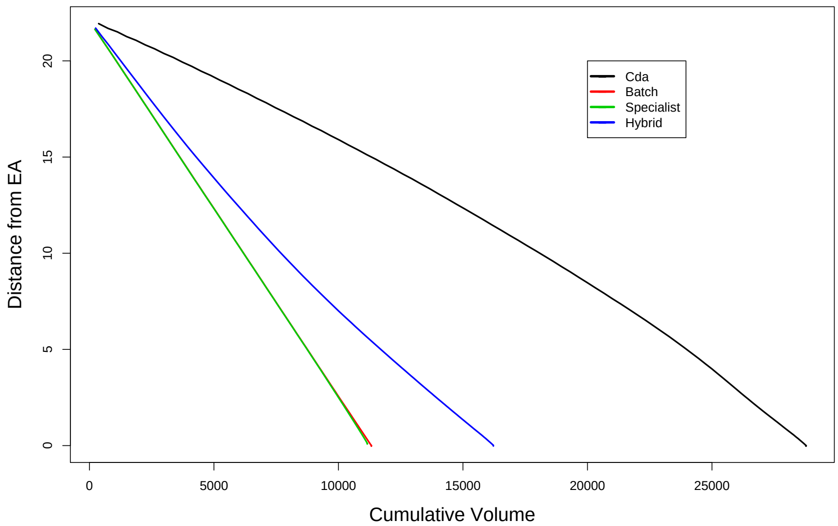

Figure 4: TT: distance from efficient allocation versus volume.

of wasteful trades during its search for allocative efficiency. In this respect, the ranking is $\{B, D\} > H > C$. Instead of volume, one may also plot the distance from the efficient allocation with respect to time. Since during a trading session all agents are randomly sampled and attempt to make a transaction, this would gauge the ability of a protocol to minimize the number of unnecessary attempted trades on the part of the traders. We do not report the plot here to save space, but the ranking it shows is $\{B, D, H\} > C$. Even though the continuous double auction generates higher trade within a session, under TT this does not work towards reducing the number of unnecessary calls in later sessions.

The fifth column in Table 2 reports the arithmetic average of traders' certainty equivalents at time NT, while the sixth column gives the percentage loss with respect to the sum of certainty equivalents associated with the Walrasian allocation. This statistics is of course zero for the batch auction and the continuous double auction, because within these protocols all exchanges take place among traders who never accept disadvantageous trades. It is relevant to evaluate the loss of welfare imposed by the presence of a specialist dealer. Because of the bid-ask spread, when a trader with negative excess demand of one unit completes a transaction with a trader with positive excess demand of one unit by going through the dealer, they end up jointly losing some money to the dealer. The ranking in this respect is $\{B, C\} > H > D$, where $>$ stands for "lower monetary losses". It is confirmed from a dynamic point of view in Figure 5.

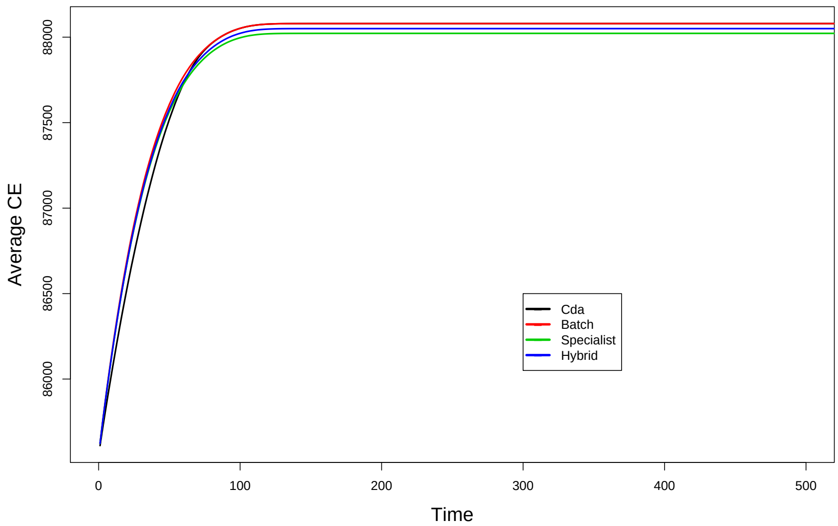

Figure 5: TT: sum of traders' certainty equivalents over time.

Finally, the eight and ninth columns in Table 2 reports the variance and the kurtosis of the time series of the closing prices from the first trading day until NT. The ranking with respect to overall volatility of prices is $H > \{B, D\} > C$, where $>$ stands for "lower volatility". Note also that both the dealership and the hybrid protocol generate empirical distributions for prices that more closely approximate the normal distribution. In a different context, LiCalzi and Pellizzari (2003) has already noted the propensity of the continuous double auction to generate non-normal statistics even under zero intelligence trading. From a dynamic point of view, Figure 6 reports the standard deviations of the closing prices computed over time windows of 20 trading days. The volatility of the continuous double

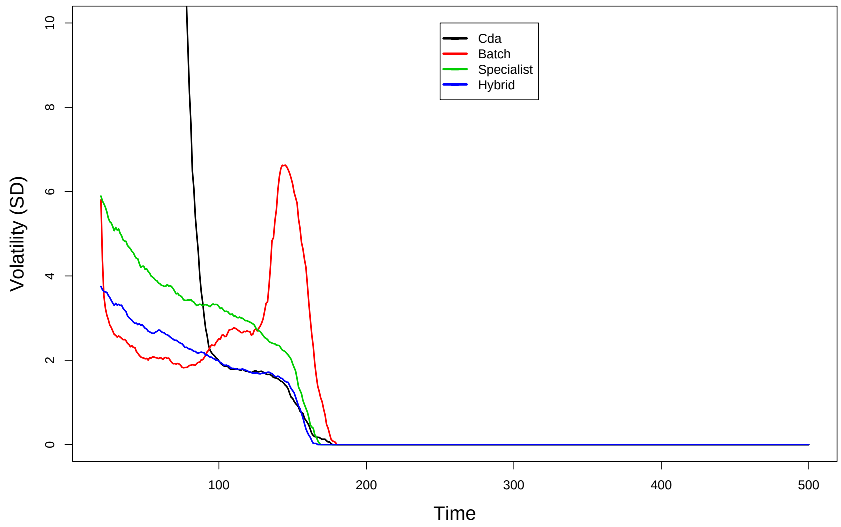

Figure 6: TT: volatillity of prices versus time. (Moving window: 20 trading days.)

auction is initially very high, but then settles down to levels similar to the hybrid protocol. The volatility of the batch auction increases when approaching the no trade time. This is a well-known consequence of the instability of the k-double auction protocol when the excess demand function has huge flats around the exchange price. We have fixed $k = 1/2$ by assumption, but the instability could be greatly reduced without affecting allocative efficiency by manipulating the value of k. Based on the dynamic comparison, the static ranking of protocols with respect to volatility should be taken with a grain of salt.

4.2 The GS assumption

Compared to TT, GS favors more efficacy over less immediacy because agents declare more extreme prices that are harder to cross. Table 3 reports summary statistics computed as averages over 25 different batches of simulations under the GS behavioral assumption. The same features described at the beginning of Section 4.1 hold.

Under GS, the rankings deduced from the table or from the dynamic comparisons (not shown here) are always the same as in TT. There is only one seeming exception from the fourth column of the table. This suggests that the ranking with respect to convergence is $H > D > C > B$, where $>$ stands for "smaller time to no trade", rather than $D > H > B > C$ as in TT. However, this difference in rankings is partly an artifact due to a

Prot.	Vol	ExcV	NT	mCE	Loss	Dist	SD	Kurt
B	11159.92	1.2%	605.28	88079.50	0.00%	0.00	1.94	42.188
C	24855.60	125%	512.96	88079.50	0.00%	0.00	16.42	22.328
D	10878.84	-0.23%	408.04	88023.01	0.064%	0.27	3.85	5.636
H	14147.56	28%	172.00	88045.47	0.039%	0.01	2.78	0.054

Table 3: GS: summary statistics for 25 passes with $n = 1000$.

few outliers. Under GS, once a protocol has reached an allocation sufficiently close to the efficient one, it has to wait until one of the few "inefficient" traders draws the right price in order to complete a transaction. In the meanwhile, several sessions go away with no trades at all. Therefore, on average full convergence under GS takes longer to be achieved. The statistics reported in the fourth column is the *average* number of trading sessions to achieve no trade and then is affected by these long waiting times.

A more stable indicator in this respect is the *median* number of trading sessions to no trade, which gives: $B(580), C(494), D(146), H(166)$. This provides two pieces of information. First, the only reversal in ranking with respect to TT concerns B and C. Second, the number of sessions needed to achieve almost full convergence under GS is comparable to what is needed for full convergence in the presence of a specialist dealer, but it is definitely larger for B and C. Intuitively, the higher efficacy sought by traders under GS slows down the attainment of efficiency only in the absence of the price-stabilizing influence of the specialist.

4.3 The MA assumption

Compared to TT, MA favors more immediacy over less efficacy because it implies that agents have less demanding requirements for agreeing to a transaction. Under TT (or GS), whenever two agents agree to a trade, their transaction must improve their joint welfare with respect to their positions immediately *before* trading. Under MA, it suffices that their joint welfare is improved with respect to their initial positions, but not necessarily with respect to their positions immediately before trading. We say that the trading under TT (or GS) is *incrementally* improving, while trading under MA is *overall* improving. This difference in the quality of the improvement sought for agreeing to a transaction probably accounts for most differences between TT (or GS) and MA.

Table 4 reports summary statistics computed as averages over 25 different batches of simulations under the MA behavioral assumption. The same features described at the beginning of Section 4.1 hold.

Prot.	Vol	ExcV	NT	mCE	Loss	Dist	SD	Kurt
B	173995.40	1478%	1361.40	88079.50	0.00%	0.00	10.92	3.473
C	248730.72	2156%	785.84	88079.47	0.00%	0.018	75.12	13.150
D	106327.36	864%	925.76	87539.94	0.61%	0.016	9.64	-0.665
H	148756.60	1248%	1081.60	87757.00	0.41%	0.01	8.76	-0.771

Table 4: MA: summary statistics for 25 passes with $n = 1000$.

Even if the MA behavioral assumption is clearly less effective towards this purpose, all protocols achieve allocative efficiency. This confirms their ability to act as partial substitutes for individual rationality. Compared to TT it now takes longer to reach the efficient allocation and the traded volumes are uniformly much higher. Traders have less demanding requirements to accept a trade and therefore much more wasteful trading goes on. The ranking with respect to volume is $D > H > B > C$, where $>$ stands for "less volume". Compared to TT, the only change in the ranking is the decline of B to third place. From a dynamic point of view, Figure 7 confirms the ranking and adds the information that the continuous double auction undergoes a sharp transition in performance. For a long initial

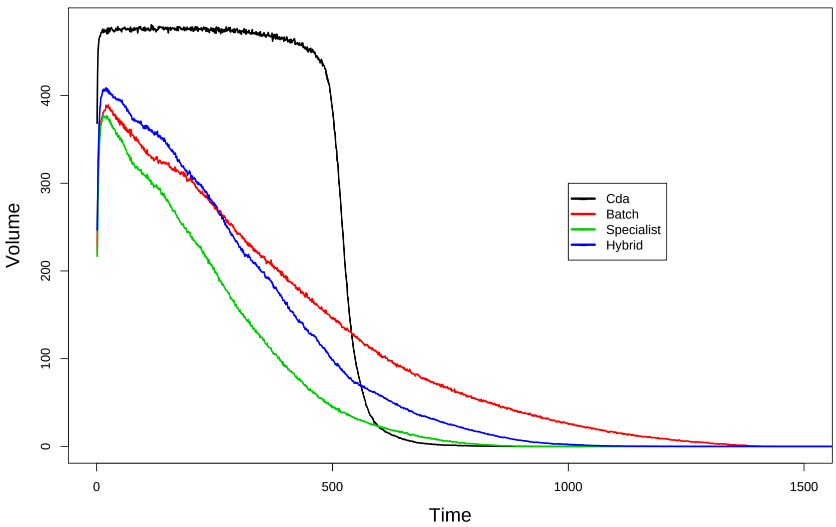

Figure 7: MA: volume versus time.

stretch of time, it generates the highest volumes. As it approaches convergence to the efficient allocation, the volume drops significantly to a minimum level, comparable only to what the dealership generates.

The ranking with respect to convergence is $C > D > H > B$, where $>$ stands for "smaller time to no trade". Compared to TT, the only change in the ranking is the improvement of C from last to first place. The dynamic analysis reveals that this increase in rank is somewhat misleading. Figure 8 shows the distance from the efficient allocation as a function of the number of trading sessions. The sharp transition in performance observed before for the continuous double auction is still there. This protocol stays farther away from the efficient allocation than the other protocols for most time, until it drops down quite abruptly (and faster) to the efficient allocation. A clearer picture emerges from Figure 9, which plots the distance against the cumulative volume and provides the ranking $D > H > B > C$ for the ability of protocols to avoid wasteful trades. Compared to TT, where the batch auction and the dealership performed similarly, the MA assumption brings out a comparative greater ability of the dealership to guide traders to the efficient allocation faster.

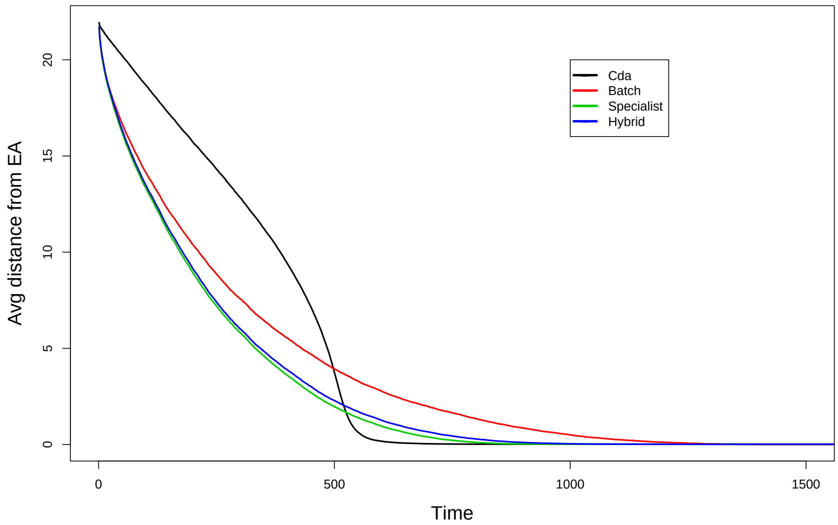

Figure 8: MA: distance from efficient allocation versus time.

The ranking with respect to the loss in welfare is the same as under TT. Consistent with the increase in traded volume, the loss of welfare imposed by the presence of a specialist dealer is now higher than under TT. Remarkably, this loss stays quite small. Figure 10 confirms the obvious ranking from a dynamic point of view.

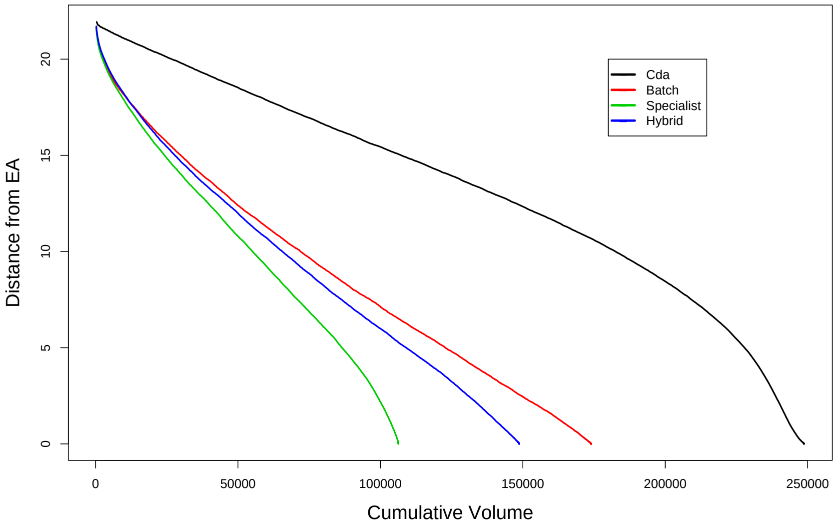

Figure 9: MA: distance from efficient allocation versus volume.

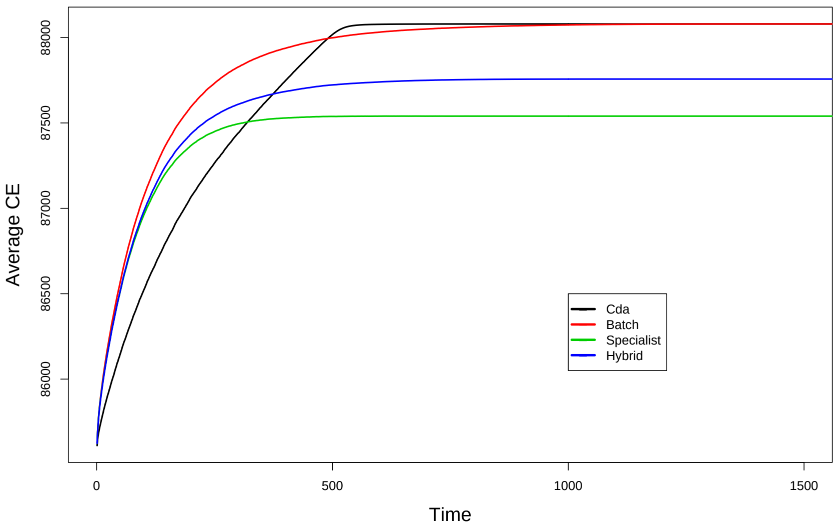

Figure 10: MA: sum of traders' certainty equivalents versus time.

The ranking with respect to overall volatility of prices is $H > D > B > C$, where $>$ stands for "lower volatility". Compared to TT, the ranking is unchanged but now unambiguously puts the batch auction behind dealership. This suggests that the nondiscretionary rule imposed on the specialist is quite effective in dampening oscillations in price, even in the mixed contexts exemplified by our hybrid protocol. From a dynamic point of view, after rescaling time, Figure 11 shows no substantial differences from the analogous plot under TT.

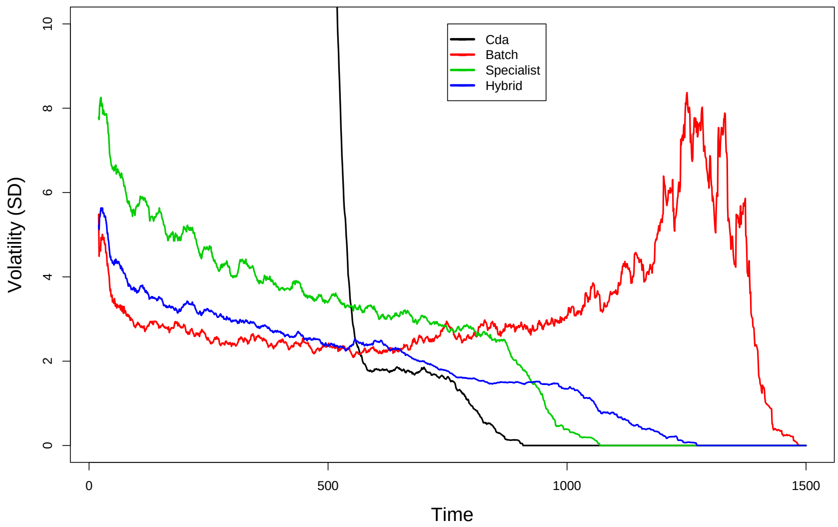

Figure 11: MA: volatillity of prices over time. (Moving window: 20 trading days.)

4.4 Tests of robustness

We have tested the robustness of our model under several different instantiations of its parameters. The only variable with a significant impact on the results is the overall liquidity of the system. When there is no sufficient cash in the system, some transactions that would enhance the allocative efficiency violate budget constraints and cannot be carried out. All the simulations reported in this paper are not affected by issues originating from insufficient liquidity.

In order to provide the reader with a more specific appreciation of this robustness, let us turn to the exemplar parametric configuration given in Table 1. We have run separate checks on each parameter, while keeping the others fixed. For each of these checks, we have obtained summary statistics (computed as averages over 25 different batches of simulations) and compared the rankings associated with them. Within a large range of values for each parameter, the rankings stay almost always unchanged. For instance, initializing $n = 250$ instead of $n = 1000$ produces only one change in the rankings: under TT, the kurtosis for B is now lower than for C. Similarly, no appreciable changes in the rankings emerge if we initialize traders' individual parameters so that exact implementation does not hold.

A particularly interesting test of robustness was run on the initial dealer's quotes. As described in Section 3.1, the exemplar configuration assumes an initial bid of 746 and an inital ask of 751. This bid-ask interval is not far from the competitive equilibrium price of 760, raising the legitimate suspicion that this might favorably bias the performance of the specialist-based protocols. Therefore, we ran two batches of 25 simulations each assuming a large variation of up to $\pm 33\%$ in the initial dealers' quotes; more precisely, we assumed an initial bid-ask interval of $[495, 501]$ and $[895, 901]$, respectively. We observed no changes in the final rankings. Clearly, when the initial interval is $[495, 501]$, the initial ask quote vastly underestimates the equilibrium price. Therefore, the dealer is initially obliged to go short on stock and match the strong incoming excess demand at prices unfavorable to her. Nonetheless, her quotes recover sufficiently fast that, when the no trade time is reached, the monetary value of her position is invariably increased; in other words, the dealer's subsequent profits from trading with a bid-ask spread suffice to make up for her initial losses. A symmetric conclusion holds when the initial quotes overestimate the equilibrium price.

5 Conclusions

The experimental literature has collected an ample empirical evidence about trading by human agents in the continuous double auction. As summarized in Smith (1982), the evidence shows that allocations and prices converge rapidly to the competitive equilibrium predictions, even if the informational requirements of this protocol are very simple. Gode and Sunder (1993) argues that the robustness of this conclusion is due to an intrinsic ability of the protocol to guide traders towards the efficient allocation. Our study confirms the allocative efficiency of the continuous double auction under two additional behavioral assumptions and, more importantly, extends Gode and Sunder's claim to other simple protocols.

Our (computerized) experiments shows that there are several simple protocols whose ability to achieve allocational efficiency seems comparable and pretty robust. Therefore, their performance should also be assessed over other relevant dimensions. Under our behavioral assumptions, a direct comparison with the batch auction shows that the continuous double action is an inferior protocol with respect to volume, time to convergence, speed to convergence, and volatility of prices. (The only exception is time to convergence under the MA assumption.) This strongly suggests that the experimental literature should give more attention to a comparative study of simultaneous versus sequential protocols; see Section 4.3.1.1 in Madhavan (2000) for a related argument.

We extended the comparison to two alternative simple protocols. The first one is a nondiscretionary form of specialist dealership, in which the rule by which quotes reacts to transactions is entirely automated and the specialist must adjust prices by one in the direction of the last trade completed. The specialist protocol is equivalent to introducing an additional agent in the market, who in a sense brings more rationality to traders' groping for the efficient allocation. However, note that the specialist is not required to exhibit zero intelligence and must accept all trades that she is proposed. It turns out that following the nondiscretionary rule suffices to produce (modest) gains while keeping her inventory under control. Intuitively, although she may occasionally lose money on some trade, our implementation of the specialist dealership improves her wealth on average. Glosten and Milgrom (1985) prove formally that this holds for a more sophisticated version of specialist dealership in an environment with diversely informed traders.

Under our behavioral assumptions, this specialist protocol has a lower time to convergence and never performs worse than the batch auction. Its only drawback is that the specialist protocol drains a tiny amount of wealth from the traders, which could on the other hand be reallocated to traders at the end of process making the specialist "neutral". This makes the specialist dealership a natural candidate for investigation in the search for effective protocols to achieve allocational efficiency.

We finally tested a hybrid protocol, that sides the specialist dealer with the continuous double auction giving each trader the option to transact at the most favorable quote available. Under our behavioral assumptions, the hybrid protocol reduces the volatility of prices and the gains of the specialist, but is otherwise inferior to the specialist protocol. The reduction in volatility and in the dealer's gains are easily explained. The specialist starts with initial quotes that may be away from the equilibrium price and must keep a fixed bid-ask spread. Almost all trades initially go through her until her quotes adjust to a level compatible with the equilibrium price. From then on, competition from the traders may occasionally provide better quotes than the dealer's bid-ask spread permits, thereby reducing volatility and her gains.

To conclude, we find that under our behavioral assumptions the four protocols generically converge to the efficient allocation in finite time. An extended comparison over other performance criteria reveals that the all-round ranking should put the specialist dealership in first place and the continuous double auction in fourth place. The exact ranking for the batch auction and the hybrid protocol depends on the weights given to the other performance criteria, although we personally judge the batch auction superior to the hybrid protocol. Finally, we remark that these conclusions hold assuming no informational effects and no strategic behavior on the part of traders.

References

[1] N. Audet, T. Gravelle and J. Yang (2002), "Alternative trading systems: Does one shoe fit all?", *Bank of Canada*, working paper 2002-33, November.

[2] G. Bottazzi, G. Dosi, and I. Rebesco (2005), "Institutional architectures and behavioral ecologies in the dynamics of financial markets", *Journal of Mathematical Economics* **41**, 197-228.

[3] D. Cliff and J. Bruten (1997), "More than zero intelligence needed for continuous double auction trading", working paper HPL-97-157, Hewlett Packard Laboratories, Bristol.

[4] L. Glosten and P. Milgrom (1985), "Bid, ask, and transaction prices in a specialist market with heterogenously informed traders," *Journal of Financial Economics* **13**, 71–100.

[5] D.K. Gode and S. Sunder (1993), "Allocative efficiency of markets with zero intelligence traders: Market as a partial substitute for individual rationality", *Journal of Political Economy* **101**, 119–137.

[6] L. Hurwicz, R. Radner and S. Reiter (1975), "A stochastic decentralized resource allocation process: Part I and Part II", *Econometrica* **43**, 187–222 and 363–394.

[7] M. LiCalzi and P. Pellizzari (2003), "Fundamentalists clashing over the book: A study of order-driven stock markets", *Quantitative Finance* **3**, 470–480.

[8] A. Madhavan (2000), "Market microstructure: A survey", *Journal of Financial Markets* **3**, 205–258.

[9] M.A. Satterthwaite and S.R. Williams (2002), "The optimality of a simple market mechanism", *Econometrica* **70**, 1841–1863.

[10] V. L. Smith (1982), "Microeconomic systems as an experimental science", *American Economic Review* **72**, 923–955.

[11] G.J. Stigler (1964), "Public regulation of the securities markets", *Journal of Business* **37**, 117–142.

[12] R. Thaler and E. Johnson (1990), "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice", *Management Science* **36**, 643–660.

[13] R. Wilson (1968), "The theory of syndicates", *Econometrica* **36**, 119–132.